Alternative Data for Credit Risk Management: An Analysis of the Current State of Research

Jan Roeder

University of Goettingen, Faculty of Business and Economics, Goettingen, Germany;
e-mail: jan.roeder@uni-goettingen.de

Abstract Determining credit risk is important for banks and non-banks alike. For credit risk management, the heterogeneous data generated today can potentially complement the established data such as balance sheet ratios. It has not yet been clearly shown which alternative data sources, such as social media or satellite data, provide added value and how this value can be extracted effectively. This review provides an overview of the intersection between these areas and develops a research agenda. The analysis of the 29 identified papers shows that the use of financial news is analyzed most frequently. Social media has also been used to some extent. The use of other alternative data sets, such as geospatial data, has been analyzed infrequently. The empirical evidence suggests that alternative data can provide both explanatory and predictive benefits in credit risk management. Convergence in terms of analytical approaches and evaluation offers the potential to advance the field.

Keywords: credit risk, alternative data, literature review, unstructured data

University of Maribor Press

DOI https://doi.org/10.18690/978-961-286-485-9.13
ISBN 978-961-286-485-9

1 Introduction

Determining credit risk is an essential task for banks and non-banks. For example, non-banks need to monitor their accounts receivables exposure and banks judge the creditworthiness of borrowers (Koulafetis 2017). In addition to established credit risk indicators (i.e., balance sheet ratios or market-based indicators (Altman and Saunders 1997)), the interest in using supplementary alternative data sources (ranging from sensor to social media data) has increased. The promising aspect is a potentially low rivalry of such data for risk management, i.e., the usefulness does not necessarily diminish with increasing dissemination of the data (Monk et al. 2019). Additionally, Mengelkamp et al. (2015) call for a further investigation of how user-generated content, which can be understood as a type of alternative data, can be integrated into corporate credit risk analysis. Although several literature reviews are situated at the intersection of alternative data (specifically text data) and finance research in general (Loughran and McDonald 2016; Nassirtoussi et al. 2014; Xing et al. 2017), there is none for the intersection of *credit risk* and *alternative data* specifically. Therefore, this paper's underlying research questions (RQ) are: 1) *What is the current state of research on using alternative data for supporting credit risk management?* And building on that: 2) *What are research gaps that future research should address?* After outlining the theoretical foundations, I define relevant parameters for the literature review. Afterwards, the results are presented and analyzed. Based on the gained insights, a research agenda is formulated, which can help to guide future research.

2 Theoretical Background

2.1 Alternative and Heterogeneous Data

Alternative data describes potentially decision-relevant but underutilized data sources, which are only available in unstructured form and cannot be used in established forecasting or risk models without prior processing (Monk et al. 2019). The potential originates from the idea that these data sources can contain important signals, for example, to identify changed customer behavior or risk situations. Especially because social media posts or anonymized credit card transactions occur with a higher frequency compared to more traditional information sources like earnings conference calls, these data sources could help to improve our risk understanding (Monk et al. 2019). The spectrum of alternative data ranges from app usage data, anonymized credit card transactions, point of sales data, or job advertisements to data on the utilization of cruise ships (alternativedata.org 2020). Roughly classified, such data can originate from individual processes, business processes, or even sensors (Monk et al. 2019). However, the universe of alternative data is so broad that no exhaustive enumeration can be provided. It should also be noted that alternative data is ultimately a collective term. There may be differences in the volume, granularity, relationality, or accuracy of the data (Monk et al. 2019; Roeder et al. 2020). This also means that different techniques for processing, storage, and analysis may be necessary.

2.2 Credit Risk Management

Credit risk describes the threat that a borrower does not repay a granted loan or fails to meet contractual obligations (Caouette et al. 2008). Types of financial risk besides credit risk include strategic risk, market risk, or compliance risk (Lam 2014). The credit risk management (CrRM) process includes the identification of risk, credit risk assessment (CRA), treatment of risk, and implementation of actions (Van Gestel and Baesens 2008). In CRA, a distinction can be made between accounting-based models (e.g., Z-score by Altman (1968)) and market-based models (Das et al. 2009). Empirical findings indicate that both approaches can be combined to increase the explanatory power (Das et al. 2009). Since the second Basel accord permits banks to use internal-rating-based approaches, they use more advanced methods (McNeil et al. 2015). Machine learning models have become increasingly relevant for credit risk prediction in recent years, of which Chen et al. (2016) provides an overview. A fundamental distinction is made based on the analyzed entity. Publicly traded companies follow strict requirements regarding corporate disclosure. Hence, accounting ratios, market-based metrics (e.g., credit default swaps (CDS), bond spreads), and credit ratings can be utilized. For private companies, the available data universe is more limited. It includes credit ratings, past transactions, industry specifics or the quality of management (Schumann 2002). However, this paper does not analyze private individuals. In a broader sense, credit risk could include all literature on stock price forecasting, etc. For this *broad* perspective, please refer to existing research (Loughran and McDonald 2016; Nassirtoussi et al. 2014; Xing et al. 2017). In contrast, this literature review focuses on research dealing with the *credit risk* of public and private companies using the automated analysis of *alternative data*. Table summarizes the scope of this paper.

Table 1: Scope of the literature review (bold is examined in more detail in this paper)

Broadness of (credit) risk understanding	Risk management	**Credit risk management**	Credit risk assessment	
Entity assessed	**Public comp.**	**Private comp.**	Individual	Sovereign

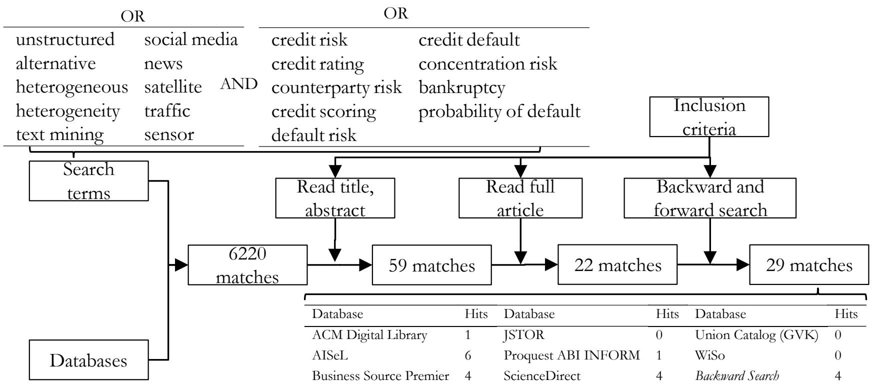
3 Research Method

This study builds on the work on literature reviews by Webster and Watson (2002) and vom Brocke et al. (2009) to synthesize existing research on alternative data for CRA. First, I define criteria for the relevance of papers. Table shows the inclusion criteria. For research to be relevant, it must examine the intersection between CrRM and alternative data. The focus lies on underexplored data sources such as news, social media, or even sensor data.

Table 2: Relevance criteria applied in the literature search

Required inclusion condition	Deals with CrRM (e.g., identifies, assesses, or monitors credit risk) of companies (public and private).
Label "Relevant"	Uses alternative data (e.g., satellite images, …) that occurs irregularly/with high frequency, uses innovative approaches.
Label "Borderline"	Uses established data sources (e.g., form 10-K data) but utilizes interesting approaches transferable to alternative data.

A comprehensive database search is conducted to identify relevant prior work. Conference and journal papers are included since the research area under investigation is still quite young. The search process is shown in Figure.



Database	Hits	Database	Hits	Database	Hits
ACM Digital Library	1	JSTOR	0	Union Catalog (GVK)	0
AISeL	6	Proquest ABI INFORM	1	WiSo	0
Business Source Premier	4	ScienceDirect	4	*Backward Search*	4
Emerald Insight	2	Springer Link	2	*Forward Search*	3
IEEE Xplore	2				

Figure 1: Literature search process

4 Results

4.1 Result Analysis

The papers in **Error! Reference source not found.** show the following **discipline** distribution (not shown in table): Finance (11), information systems (12), and computer science (6). The moderate number of papers can be attributed to the novelty of this research area. The availability of data sources is tightly coupled with the **entity** whose credit risk is **assessed**. As a side note, papers tend to exclude banks because their financial metrics are distributed differently and therefore may not be comparable (e.g., Lu et al. (2015)). The number of papers analyzing public companies is distributed relatively evenly between non-banks (27) and banks (19). Not publicly traded companies were analyzed less frequently (6). The reason could be that more data is available for exchange-listed companies due to the disclosure obligations. Regarding the **granularity** of the analysis, eight studies are carried out at the annual level and five at the quarterly level. In these studies, accounting ratios can be integrated into the analysis straightforwardly. The situation is different for the analyses at a monthly (1), weekly (2), and daily level (4), where traditional regression analyses can be problematic due to autocorrelation issues when using quarterly numbers. Regarding alternative **data sources**, it is apparent that financial news (11) has been analyzed the most, followed by annual reports including 10-K filings (8) and posts on social media (6). Other data sources are transactions on B2B platforms

(1), search engine sentiment (1), or data on staffing decisions (1). No research on using textual financial analyst reports for CrRM could be identified. For the **dependent variables**, a categorical variable is used eight times and the credit rating by rating agencies seven times. The CDS spread is also used seven times as a dependent variable. In addition, the bond spread (2), the LIBOR-OIS spread (1), and equity volatility (1) need to be mentioned. One difference between the two main categorical measures and the CDS spread is that the CDS spread is a continuously updated market-based measure.

Error! Reference source not found. also shows the **statistical models**, which link the endogenous variable to the signals extracted from alternative data sources. Finance-centric literature often uses logistic, linear, and panel regression to ensure sufficient interpretability. Here, the clustering of the data regarding the company and time period tends to be modeled explicitly, as is done using panel regression models in seven cases (e.g., Bao and Datta (2014); Liebmann et al. (2016); Tsai et al. (2016)). Contributions focused on machine learning often use more complex models (e.g., random forest or neural network) and optimize them extensively to optimize the predictive accuracy. Papers also develop custom architectures. For example, Zhao et al. (2019) propose a network architecture to incorporate financial variables and unstructured data. The **data split** (not in **Error! Reference source not found.** due to space limitations) indicates the extent to which a model's forecasting capability has been verified. Fundamentally, both traditional statistical approaches (hypothesis testing and regression), often assessed in-sample, as well as insights stemming from machine learning, which often uses cross-validation, make important contributions. The identified data splits range from classical regression without a split (15) (Bao and Datta 2012; Hu et al. 2018) over the two-way split (10), i.e., train-test split, (Altman et al. 2010) to cross-validation (2), e.g., (Choi et al. 2020) and not available (2). Due to the relatively large number of publications without a data split, statements regarding the out-of-sample performance need to be interpreted cautiously.

Regarding the central **findings** of the identified papers, the importance of the sentiment extracted from financial news for CRA could be shown quite consistently in varying research setups (Janner and Schmidt 2015; Liebmann et al. 2016; Lu et al. 2012; Norden 2017). The same applies to the volume of news (Tsai et al. 2016). The results are more ambiguous for automated topic extraction since the topic models and assigned topic labels vary from study to study. (Bao and Datta 2014; Hajek et al.

2017). For example, the topics "restructuring" and "investment policy" are found to be important in one study (Hajek et al. 2017) and "macroeconomic risk" and "funding risk" in another (Bao and Datta 2014). Fernandes and Artes (2016) and González-Fernández and González-Velasco (2019) represent more unusual approaches. The former uses spatial data to improve the CRA; the latter shows that search engine activity correlates with a credit risk measure. All in all, most papers identify an added value of the variables obtained from alternative data, providing explanatory value or improving the predictive power. Concerning design knowledge, few papers are fully situated in the design science research (DSR) paradigm. Design requirements or principles remain rather implicit. Hristova et al. (2017) propose the *RatingBot* and a process to extract a credit rating from text but the requirements arising from the problem domain could be formulated more explicitly. Zhao et al. (2019) propose a default prediction framework. However, the abstraction towards design principles or even theories is absent. Hence, a lot of untapped potential for research in the DSR paradigm is apparent.

Table 3: Publications identified in the literature review
(If not specified "Entities" includes banks and non-banks; * signals borderline relevance)

Research paper	**Entities Granul.**	**Alt. data source**	**Risk measure (dep. variable)**	**Statistical Model**	**Main results**
(Aktug et al. 2015)	PU E	Human resource	Bond spread	Event study, hypothesis test	Credit analyst hiring has an impact on bond but not stock return
(Altman et al. 2010)	PR NB Q	Other	Default category	Log. reg.	Alternative data on legal actions by creditors helps to increase predictive power
(Bao and Datta 2012)*	PU Y	Annual reports	Unsupervised	Not applicable	Proposed Sent-LDA improves identification of risk types
(Bao and Datta 2014)*	PU Y	Annual reports	Equity volatility	Panel reg.	With Sent-LDA, 2/3 risk types not relevant, three show risk increase and five decrease
(Cecchini et al. 2010)*	PU Y	Annual reports	Default category	SVM	Fin. variables with complementary qualitative data achieves the best result
(Choi et al. 2020)*	PU NB Y	Annual reports	Credit rating	SVM, NN, RF	Combination of fin. variables and qualitative data achieves the best prediction

(Fernandes and Artes 2016)	PR NB NA	Spatial data	Default category	Log. reg.	The inclusion of a spatial risk factor improved bankruptcy identification
(González-Fernández and González-Velasco 2019)	PU BA W	Search engine	CDS spread	Linear reg.	Inclusion of search engine-based sentiment index improved credit risk prediction
(Gül et al. 2018)	PU NA	Social media	Credit rating	Multiple criteria decision making	Social media was found to be useful in CRA for half of the companies
(Hajek et al. 2017)*	PU Y	Annual reports	Credit rating	Log. reg., decision tree, NBN, RF, …	Risky firms mention restructuring less and domestic market difficulties more frequently
(Hristova et al. 2017)*	PU Y	Annual reports	Credit rating	Log reg., naïve bayes, SVM, …	Classification model and text representation are important determinants of accuracy
(Hu et al. 2018)*	PU Y	Annual reports	CDS spread	Panel reg.	Evidence for an inverse relationship between readability and CDS spreads
(Janner and Schmidt 2015)	PU E	Financial news	Bond spread	Event study hypothesis test	Explanatory power of corp. news for bonds is comparable to power for stock market
(Liebmann et al. 2016)	PU D	Financial news	CDS spread	Panel reg.	Based on sentiment, CDS traders and equity traders interpret the same news differently
(Lu et al. 2012)	PU NB Q	Financial news	Credit rating	Mod. probit, SVM	News coverage is significantly associated with credit downgrades
(Lu et al. 2015)	PU NB Q	Financial news	Default category	Log. reg.	Distress indicator derived from financial news possesses significant explanatory power
(Mengelkamp et al. 2015)	PU,PR NB M	Social media	Default category	Hypothesis test, k-nearest-neighbor	More social media posts and worse sentiment for financially instable companies; classification accuracy above 50%.
(Mengelkamp et al. 2016)	PU,PR NB NA	Social media	Default category	Frequency counts	Sentiment dictionary achieves 67.9% accuracy compared to 49.97% by domain-independent dictionary

(Mengelkamp et al. 2017)*	PU,PR NB NA	Social media	Default category (proxy)	Log. reg., dec. tree, SVM, …	SVM and bag-of-words show the best performance for Tweet classification
(Norden 2017)	PU D	Financial news	CDS spread	Panel reg.	Financial news show significant influence on the way CDS spreads change
(Onay and Öztürk 2018)	PU NA	Social media	Not applicable	Not applicable	Review shows rising relevance of non-traditional data sources for credit scoring
(Safi and Lin 2014)	PR NB NA	Commerce platform	Solvency proxy	Log. reg.	Measures from commerce platform (membership period, page views) help to explain creditworthiness
(Smales 2016)	PU BA D	Financial news	CDS spread, LIBOR spread	Panel reg.	Significant negative relationship between news sentiment and CDS spread changes
(Tsai et al. 2010)	PU Y	Financial news	Credit rating	Ordered logit/ probit model	Sentiment analysis of corporate news shows explanatory contribution for credit rating
(Tsai et al. 2016)	PU NB Q	Financial news, Ann. reports	CDS spread	Panel reg.	High news volume and negative sentiment are associated with an increase in credit risk
(Yan et al. 2019)	PU NA	Financial news	Entity association	Uni- and bidirect. GRU	Modeling of relation between firms using neural network improves classification
(Yang et al. 2020)	PU D	Financial new	CDS spread	Panel reg.	Inverse relationship of news sentiment and CDS spread; more pronounced in cases of higher analysts' earnings forecasts dispersion
(Yuan et al. 2018)	PU Q	Social media	Credit rating	Log reg., RF, NN, SVM	Topic model that incorporates emotion detection achieves improved accuracy
(Zhao et al. 2019)	PU W	Financial news	Default category	GAM, NN	Combination of financial measures and social media data improves accuracy

Table 4: Definition of abbreviations

Entities	PU (public), PR (private), BA (bank), NB (non-bank)
Granularity	E (Events), D (Daily); W (Weekly), M (Monthly), Q (Quarterly), Y (Yearly), NA (not available)
Stat. model	Logistic regression (log. reg.), panel regression (panel reg.), support vector machine (SVM), neural network (NN), random forest (RF), naïve Bayesian network (NBN), gated recurrent unit (GRU), generalized additive model (GAM)

4.2 Research Agenda

The following **research agenda for future research** on **using alternative data for CrRM** is derived from the analysis. Research gap (*RG*) *1. research on non-public companies*. There is still a lack of research that focuses on small and medium-sized non-public companies since there is less information (e.g., market-based metrics) available. *RG 2. need for research using irregularly occurring/frequent data sources*. There is still untapped potential to analyze signals from alternative data with varying frequency to support decision-making in CrRM. *RG 3. research beyond quarterly frequency*. For the used methods and findings, more research on how alternative data can support CrRM is needed, ranging from the monthly to intraday level. *RG 4. convergence between econometrics and machine learning-based studies regarding methods and evaluation*. There is a distinct divergence between the fields, which calls for more interdisciplinary research to allow rigorous evaluations and more comparable results. *RG 5. Research on design knowledge*. Due to the rather implicit use of DSR in many cases, further research is needed to create expand knowledge related to the design of risk management systems utilizing alternative data (principles and theories).

5 Discussion and Conclusion

In terms of *practical implications*, the studies suggest that alternative data can contain decision-relevant signals. The question arises how the CrRM process and skill profiles (data integration/analysis) for risk managers may evolve. Will the in-house capabilities for CRA need to increase, or will such insights be procured from external entities? This paper *contributes* to *research* by 1) identifying and classifying research addressing the intersection of CrRM and alternative data and 2) deriving a research agenda that encompasses the most prominent research gaps. Like any research, this study has potential *limitations*. Literature reviews are affected by publication bias.

Additionally, the search could fail to identify relevant literature because related terms were not considered. Also, the alternative data that studies use was mostly in English. Moreover, only published research could be considered, not proprietary models used by banks. To answer RQ1, the literature review classified 29 papers in which alternative data provides a basis for enhanced CrRM. The proposed research agenda consists of the most prominent research gaps, thereby addressing RQ2. Overall, the evidence suggests that alternative data can improve CrRM in terms of a better understanding of the risk situation and the predictive performance.

References

Aktug, R. E., Nayar, N., and Salas, J. M. 2015. "Are Credit Rating Agency Analysts Valuable?," The Journal of Risk Finance (16:4), pp. 378-394.

alternativedata.org. 2020. Retrieved 02/25/2021, from www.alternativedata.org.

Altman, E. I. 1968. "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," The Journal of Finance (23:4), pp. 589-609.

Altman, E. I., Sabato, G., and Wilson, N. 2010. "The Value of Non-Financial Information in Small and Medium-Sized Enterprise Risk Management," Journal of Credit Risk (6:2), pp. 1-33.

Altman, E. I., and Saunders, A. 1997. "Credit Risk Measurement: Developments over the Last 20 Years," Journal of Banking & Finance (21:11-12), pp. 1721-1742.

Bao, Y., and Datta, A. 2012. "Summarization of Corporate Risk Factor Disclosure through Topic Modeling," Proceedings of the 33rd International Conference on Information Systems, Orlando, FL.

Bao, Y., and Datta, A. 2014. "Simultaneously Discovering and Quantifying Risk Types from Textual Risk Disclosures," Management Science (60:6), pp. 1371-1391.

Caouette, J. B., Altman, E. I., Narayanan, P., and Nimmo, R. 2008. Managing Credit Risk: The Next Great Financial Challenge. Hoboken, New Jersey: John Wiley & Sons, Inc.

Cecchini, M., Aytug, H., Koehler, G. J., and Pathak, P. 2010. "Making Words Work: Using Financial Text as a Predictor of Financial Events," Decision Support Systems (50:1), pp. 164-175.

Chen, N., Ribeiro, B., and Chen, A. 2016. "Financial Credit Risk Assessment: A Recent Review," Artificial Intelligence Review (45:1), pp. 1-23.

Choi, J., Suh, Y., and Jung, N. 2020. "Predicting Corporate Credit Rating Based on Qualitative Information of Md&a Transformed Using Document Vectorization Techniques," Data Technologies and Applications (54:2), pp. 151-168.

Das, S. R., Hanouna, P., and Sarin, A. 2009. "Accounting-Based Versus Market-Based Cross-Sectional Models of Cds Spreads," Journal of Banking & Finance (33:4), pp. 719-730.

Fernandes, G. B., and Artes, R. 2016. "Spatial Dependence in Credit Risk and Its Improvement in Credit Scoring," European Journal of Operational Research (249:2), pp. 517-524.

González-Fernández, M., and González-Velasco, C. 2019. "An Alternative Approach to Predicting Bank Credit Risk in Europe with Google Data," Finance Research Letters (35).

Gül, S., Kabak, Ö., and Topcu, I. 2018. "A Multiple Criteria Credit Rating Approach Utilizing Social Media Data," Data & Knowledge Engineering (116), pp. 80-99.

Hajek, P., Olej, V., and Prochazka, O. 2017. "Predicting Corporate Credit Ratings Using Content Analysis of Annual Reports – a Naïve Bayesian Network Approach," FinanceCom 2016, Frankfurt, Germany: Springer International Publishing, pp. 47-61.

Hristova, D., Probst, J., and Eckrich, E. 2017. "Ratingbot: A Text Mining Based Rating Approach," Proceedings of the 38th Int. Conference on Information Systems, Seoul, South Korea, pp. 1-19.

Hu, N., Liu, L., and Zhu, L. 2018. "Credit Default Swap Spreads and Annual Report Readability," Review of Quantitative Finance and Accounting (50:2), pp. 591-621.

Janner, S., and Schmidt, D. 2015. "Are Economically Significant Bond Returns Explained by Corporate News? An Examination of the German Corporate Bond Market," Financial Markets and Portfolio Management (29:3), pp. 271-298.

Koulafetis, P. 2017. Modern Credit Risk Management: Theory and Practice. London: Palgrave Macmillan.

Lam, J. 2014. Enterprise Risk Management: From Incentives to Controls. Hoboken, New Jersey: John Wiley & Sons, Inc.

Liebmann, M., Orlov, A. G., and Neumann, D. 2016. "The Tone of Financial News and the Perceptions of Stock and Cds Traders," Int. Review of Financial Analysis (46), pp. 159-175.

Loughran, T., and McDonald, B. 2016. "Textual Analysis in Accounting and Finance: A Survey," Journal of Accounting Research (54:4), pp. 1187-1230.

Lu, H.-M., Tsai, F.-T., Chen, H., Hung, M.-W., and Li, S.-H. 2012. "Credit Rating Change Modeling Using News and Financial Ratios," ACM Trans. on Management Inf. Systems (3:3), pp. 1-30.

Lu, Y.-C., Wei, Y.-C., and Chang, T.-Y. 2015. "The Effects and Applicability of Financial Media Reports on Corporate Default Ratings," Int. Review of Economics & Finance (36), pp. 69-87.

McNeil, A. J., Frey, R., and Embrechts, P. 2015. Quantitative Risk Management: Concepts, Techniques and Tools - Revised Edition, (2 ed.). Princeton, New Jersey: Princeton University Press.

Mengelkamp, A., Hobert, S., and Schumann, M. 2015. "Corporate Credit Risk Analysis Utilizing Textual User Generated Content - a Twitter Based Feasibility Study," Proceedings of the 19th Pacific Asia Conference on Information Systems, Singapore.

Mengelkamp, A., Rojahn, M., and Schumann, M. 2017. "Evaluating Machine Learning Algorithms for Sentiment Classification of Tweets in Credit Assessment," ECSM 2017 4th European Conference on Social Media, Vilnius, Lithuania, pp. 209-216.

Mengelkamp, A., Wolf, S., and Schumann, M. 2016. "Data Driven Creation of Sentiment Dictionaries for Corporate Credit Risk Analysis," Proceedings of the 22nd Americas Conference on Information Systems, San Diego, CA.

Monk, A., Prins, M., and Rook, D. 2019. "Rethinking Alternative Data in Institutional Investment," The Journal of Financial Data Science (Winter 2019), pp. 14-31.

Nassirtoussi, A. K., Aghabozorgi, S., Ying Wah, T., and Ngo, D. C. L. 2014. "Text Mining for Market Prediction: A Systematic Review," Expert Systems with Applications (41:16), pp. 7653-7670.

Norden, L. 2017. "Information in Cds Spreads," Journal of Banking & Finance (75), pp. 118-135.

Onay, C., and Öztürk, E. 2018. "A Review of Credit Scoring Research in the Age of Big Data," Journal of Financial Regulation and Compliance (26:3), pp. 382-405.

Roeder, J., Muntermann, J., and Kneib, T. 2020. "Towards a Taxonomy of Data Heterogeneity," Proceedings of the 15th International Conference on Wirtschaftsinformatik, Potsdam, Germany.

Safi, R., and Lin, Z. 2014. "Using Non-Financial Data to Assess the Creditworthiness of Businesses in Online Trade," Proceedings of the 18th Pacific Asia Conf. on Information Systems, Chengdu, China.

Schumann, M. 2002. "Bonitätsbeurteilungen Und Kreditprüfungen Im Electronic Business," in Electronic Business: Theoretische Aspekte Und Anwendungen in Der Betrieblichen Praxis, R. Gabriel and U. Hoppe (eds.). Heidelberg: Physica-Verlag HD, pp. 413-427.

Smales, L. A. 2016. "News Sentiment and Bank Credit Risk," Journal of Empirical Fin. (38), pp. 37-61.

Tsai, F.-T., Lu, H.-M., and Hung, M.-W. 2010. "The Effects of News Sentiment and Coverage on Credit Rating Analysis," Proceedings of the 14th Pacific Asia Conf. on Inf. Systems, Taipei, Taiwan.

Tsai, F.-T., Lu, H.-M., and Hung, M.-W. 2016. "The Impact of News Articles and Corporate Disclosure on Credit Risk Valuation," Journal of Banking & Finance (68), pp. 100-116.

Van Gestel, T., and Baesens, B. 2008. Credit Risk Management: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital. Oxford, United Kingdom: Oxford University Press.

vom Brocke, J., Simons, A., Niehaves, B., Niehaves, B., Reimer, K., Plattfaut, R., and Cleven, A. 2009. "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process," Proceedings of the 17th European Conference on Information Systems, Verona, Italy.

Webster, J., and Watson, R. T. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," MIS Quarterly (26:2), pp. xiii-xxiii.

Xing, F. Z., Cambria, E., and Welsch, R. E. 2017. "Natural Language Based Financial Forecasting: A Survey," Artificial Intelligence Review (50:1), pp. 49-73.

Yan, C., Fu, X., Wu, W., Lu, S., and Wu, J. 2019. "Neural Network Based Relation Extraction of Enterprises in Credit Risk Management," 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), Kyoto, Japan, pp. 1-6.

Yang, S., Liu, Z., and Wang, X. 2020. "News Sentiment, Credit Spreads, and Information Asymmetry," North American Journal of Economics & Finance (52).

Yuan, H., Lau, R. Y. K., Wong, M. C. S., and Li, C. 2018. "Mining Emotions of the Public from Social Media for Enhancing Corporate Credit Rating," 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), Xi'an, China, pp. 25-30.

Zhao, Y., Shen, Y., and Huang, Y. 2019. "Dmdp: A Dynamic Multi-Source Default Probability Prediction Framework," Data Science and Engineering (4:1), pp. 3-13.